nature research

Corresponding author(s): Mihai Netea

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	In total 534 adult volunteers from Nijmegen, the Netherlands (237 males and 296 females, age range 18–75 years) where included in 500FG. After visiting the hospital to donate blood, the volunteers received an extensive online questionnaire about lifestyle, diet, and disease history. Based on the results of this questionnaire 45 volunteers where excluded for various reasons, e.g., they were under medication, non-European ancestry, or had a chronic disease. By excluding these individuals from the analysis we minimized false positive effects on the cytokine production capacity in vitro and in vivo.
Data analysis	The analyses where performed using the statistical programming language R and several publicly available and previously published methods/library's. No custom algorithms have been developed in this study and all analyses were done using either base R 3.2 or previously published methods. Further software that was used consisted of the open source toolkits PLINK 1.9, FastQC 0.11, STAR 2.4 and SAMTools 1.4

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available at https://hfgp.bbmri.nl/ were it has been meticulously catalogued and archived at BBMRI-NL aiming for maximum reuse following the FAIR principles, i.e., Findability, Accessibility, Interoperability, and Reusability. Individual level genetic data as well as other privacy sensitive datasets are available upon request at http://www.humanfunctionalgenomics.org/site/?page_id=16. These datasets are not publicly available because they contain information that could compromise the research participants privacy. The central data stewardship and access has been implemented using MOLGENIS open source platform for scientific data that enables flexible data upload, management and querying, including sufficiently rich metadata and interfaces for machine processing and custom (R statistics) visualization for human processing (see http://molgenis.org). Also summaries of the study have been submitted to BBMRI central catalogues https://catalogue.bbmri.nl (Netherlands) and http://www. bbmri-eric.eu/news-events/bbmri-eric-directory-2-0/ (EU).

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Since no the proportion of genetic variation is unknown sample size calculation was not possible. During a previous pilot study using 79 samples multiple genome wide significant QTL's were found indicating that significant effects can be observed even with limited sample size. In the current analysis we utilized a dataset of ~ 500 individuals for increased power, which allowed us to observe more significant associations. With this we are the largest study to date on cytokine production.
Data exclusions	For several of the datasets samples where excluded during the QC process. For genotyping several samples where excluded due to being ethnic outliers. Individuals with severe disease or under heavy medication were excluded from the entire analysis as well.
Replication	Most of the experiments where replicated using statistical re-sampeling methods to ensure the reliability of the models produced. The prediction models for cytokine response were applied in an independent cohort and several of the tested models could be validated. In addition, we performed several in vitro experiments testing several associations which were found. These experiments validated our hypotheses.
Randomization	Where applicable, samples (individuals) where divided into testing an training groups randomly multiple times to ensure the quality of the models produced.
Blinding	Blinding was not applicable to the current study since no group allocation was performed. Immune variation was studied in a population based cohort where individuals were not allocated into groups during data collection.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	The main analyses were performed in the 500FG cohort, which is part of the Human Functional Genomics Project. This cohort consists of 534 healthy individuals (237 males and 296 females) of Caucasian origin. Volunteers range from 18 to 75 years of age, with the majority (421 individuals) being 30 years or younger. BMI is within normal limits (15 to 35) with the majority (380 individuals) having a BMI between 20 and 25 . Of these 534 original volunteers, 45 were excluded based on genetic background and questionnaire results (medication usage, chronic disease) leaving 489 individuals.
Recruitment	Recruitment of samples for 500FG was done on a volunteer basis. The final selection of volunteers included a bias towards younger individuals. This has a potential impact on the generality of the results observed i.e. results observed might not be directly applicable in an older age group.